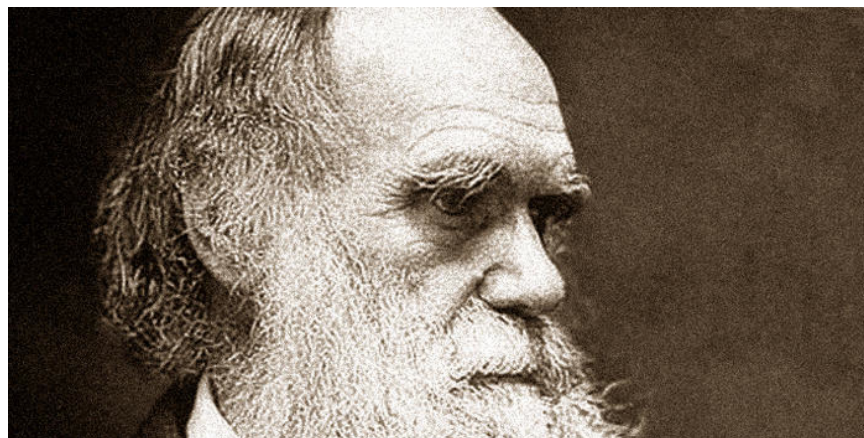# ISO 19157 – A Framework for Progress on Data Quality

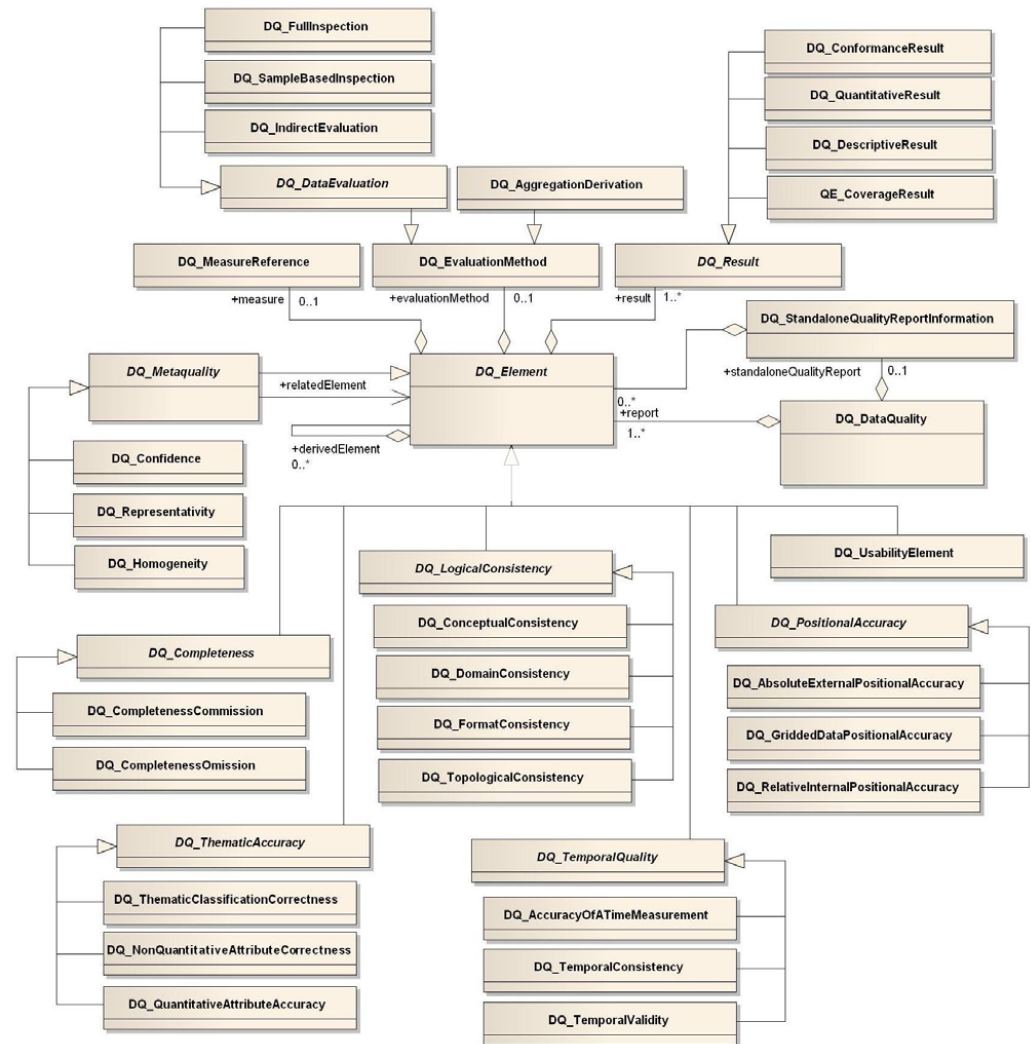The ISO Data Quality Metadata standards are evolving.

Evolution is good.

Ted Habermann

Director of Earth Science

The HDF Group

thabermann@hdfgroup.org

ISO 19157 is a conceptual model of data quality that was recently approved as an international standard. It combines concepts from three older standards into a unified conceptual model for describing data quality.

Many of the principle elements of this conceptual model are abstract, and can be implemented in several ways.

The Big Picture

ISO 19157 is a conceptual model of data quality that was recently approved as an international standard. It combines concepts from three older standards into a unified conceptual model for describing data quality.
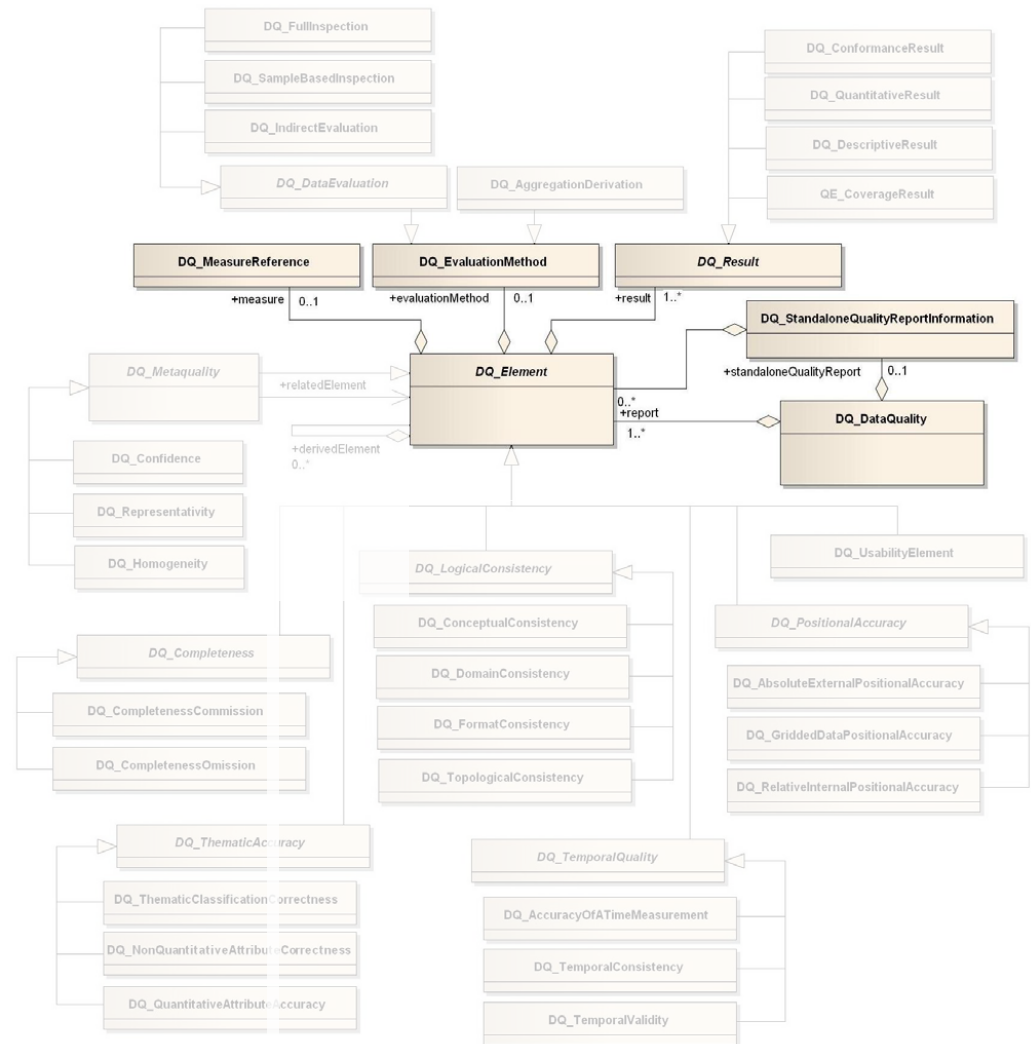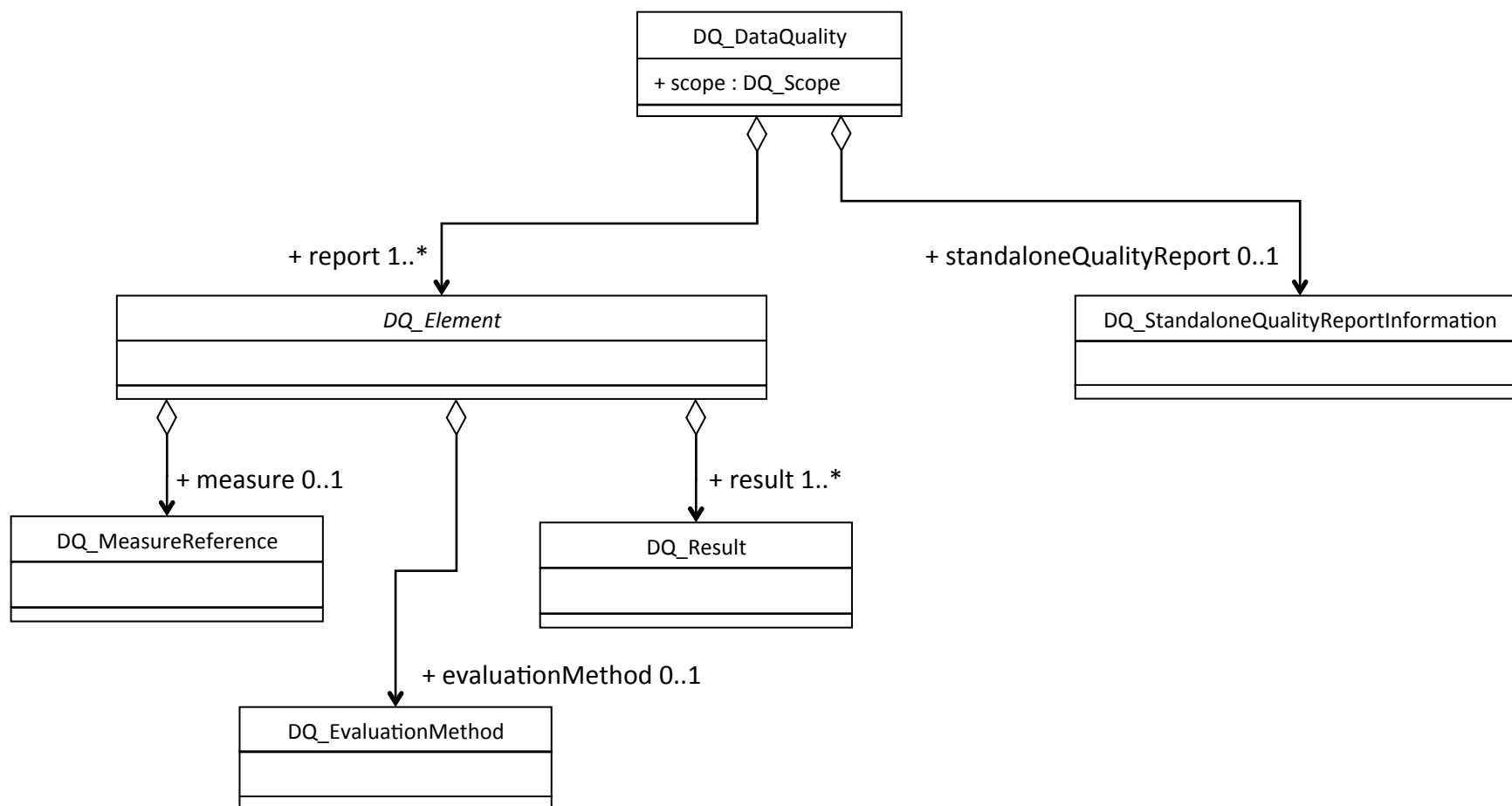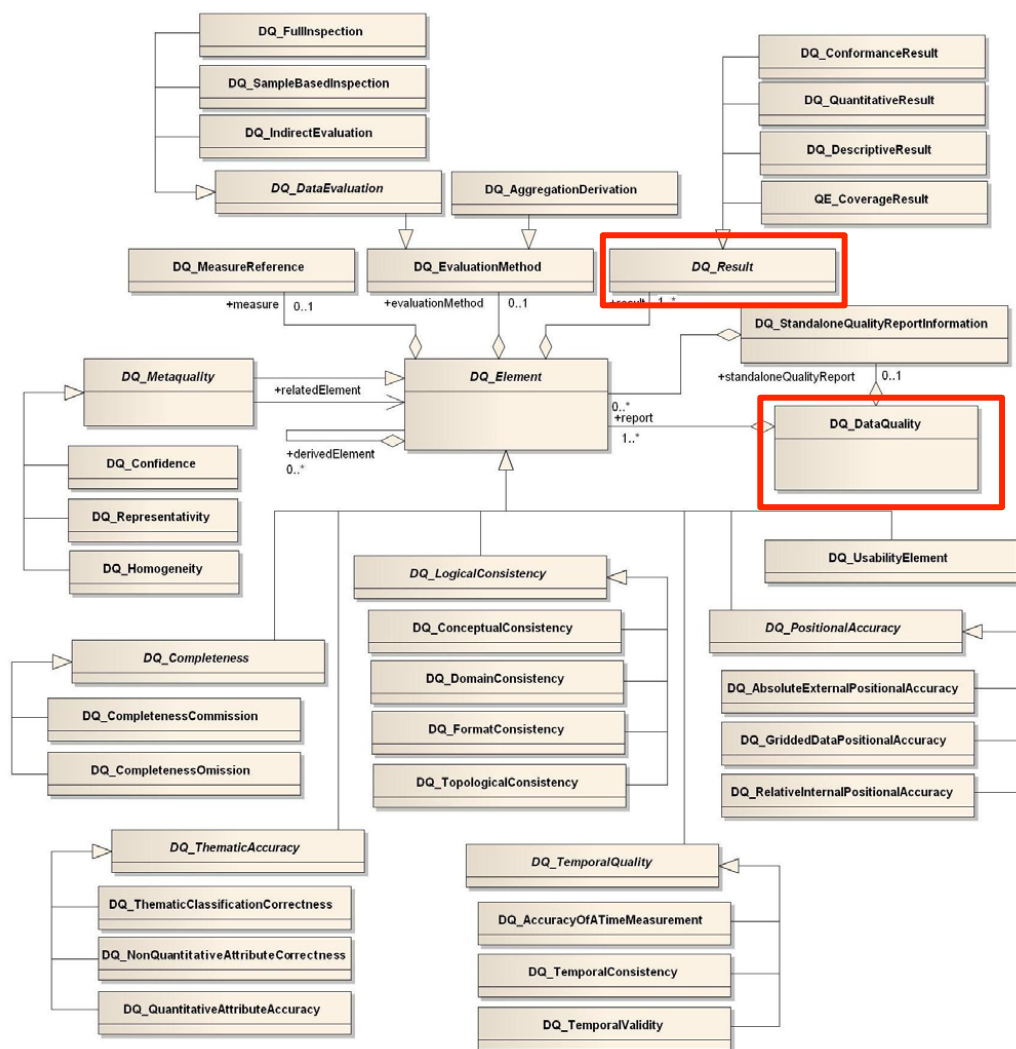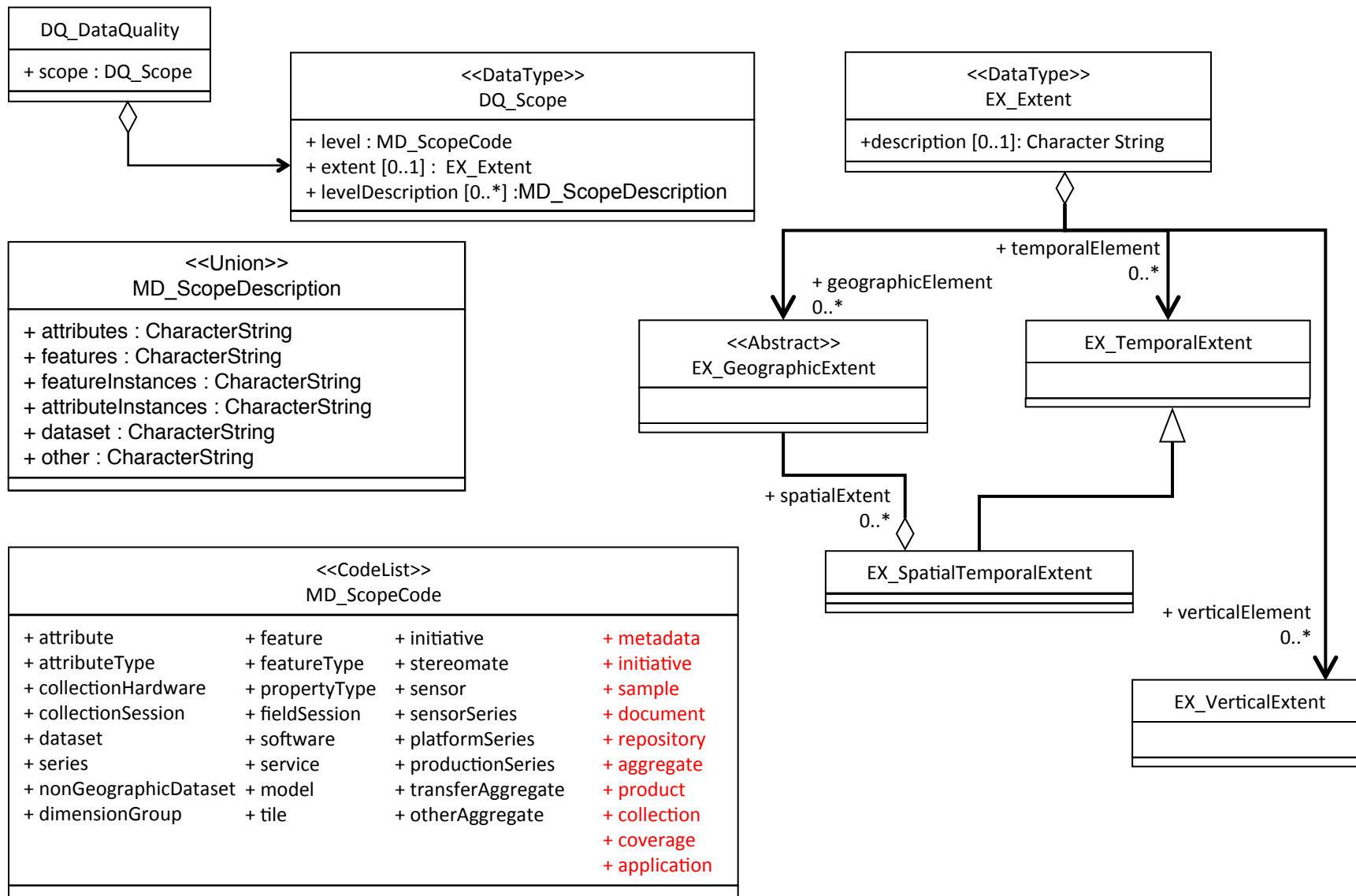
Many of the principle elements of this conceptual model are abstract, they can be implemented in several ways.
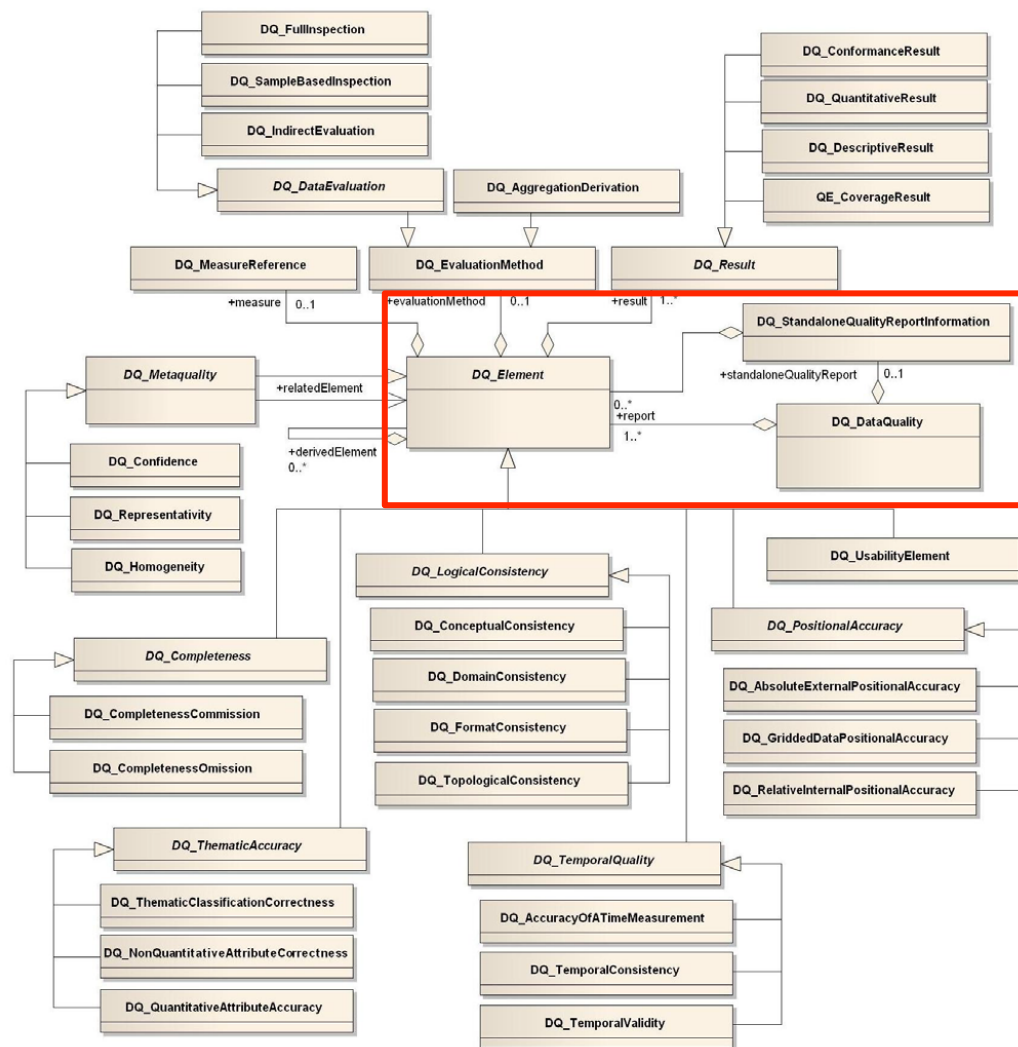
When only the abstract concepts are considered, the model is very simple.

DQ_DataQuality

+ scope : DQ_Scope

<<DataType>>
DQ_Scope

+ level : MD_ScopeCode
+ extent [0..1] : EX_Extent
+ levelDescription [0..*] :MD_ScopeDescription

<<DataType>>
EX_Extent

+description [0..1]: Character String

<<Union>>
MD_ScopeDescription

+ attributes : CharacterString
+ features : CharacterString
+ featureInstances : CharacterString
+ attributeInstances : CharacterString
+ dataset : CharacterString
+ other : CharacterString

+ temporalElement
0..*

+ geographicElement
0..*

<<Abstract>>
EX_GeographicExtent

EX_TemporalExtent

+ spatialExtent
0..*

EX_SpatialTemporalExtent

+ verticalElement
0..*

EX_VerticalExtent

<<CodeList>>
MD_ScopeCode

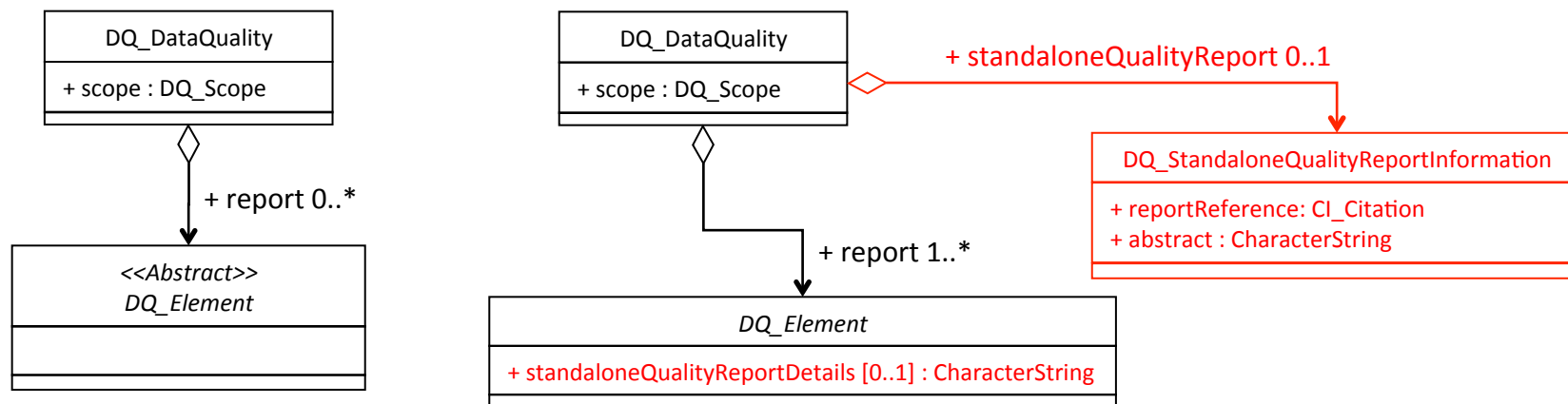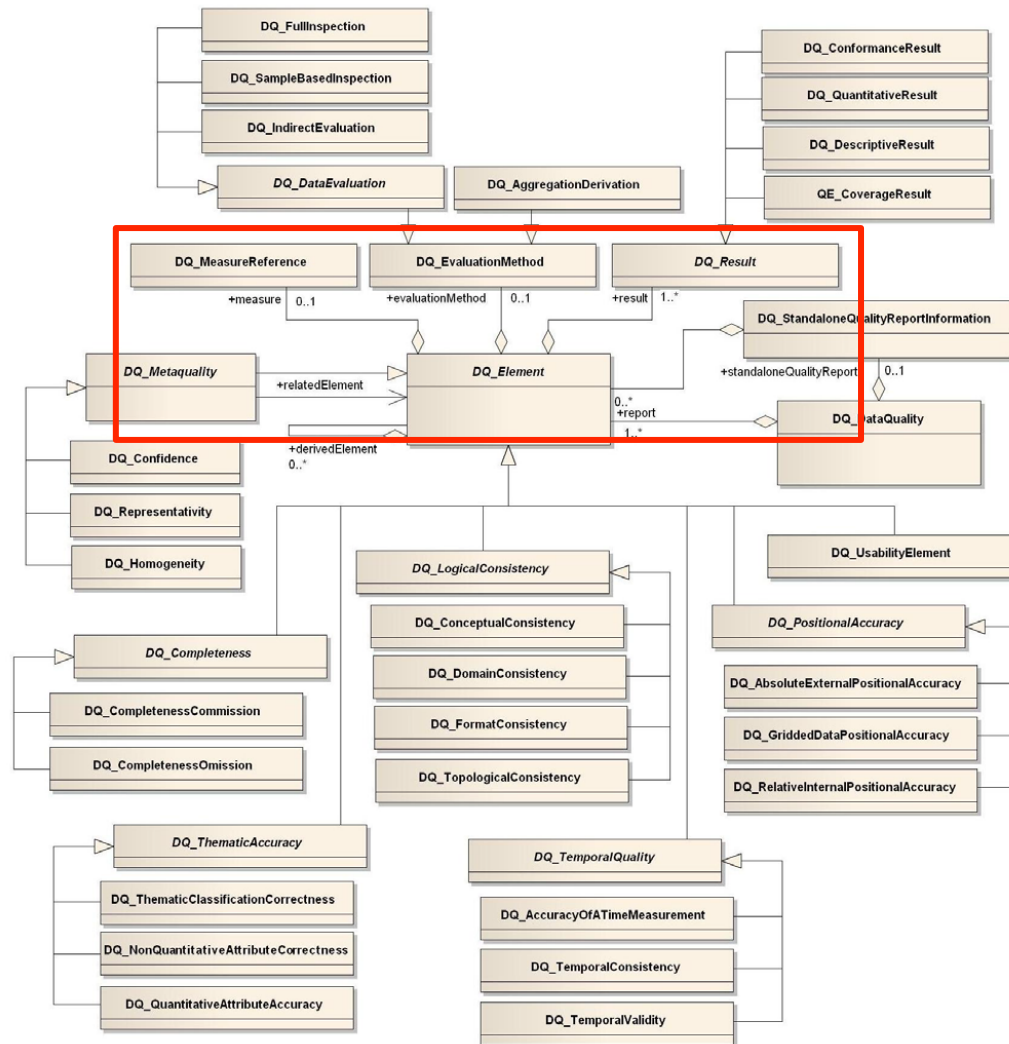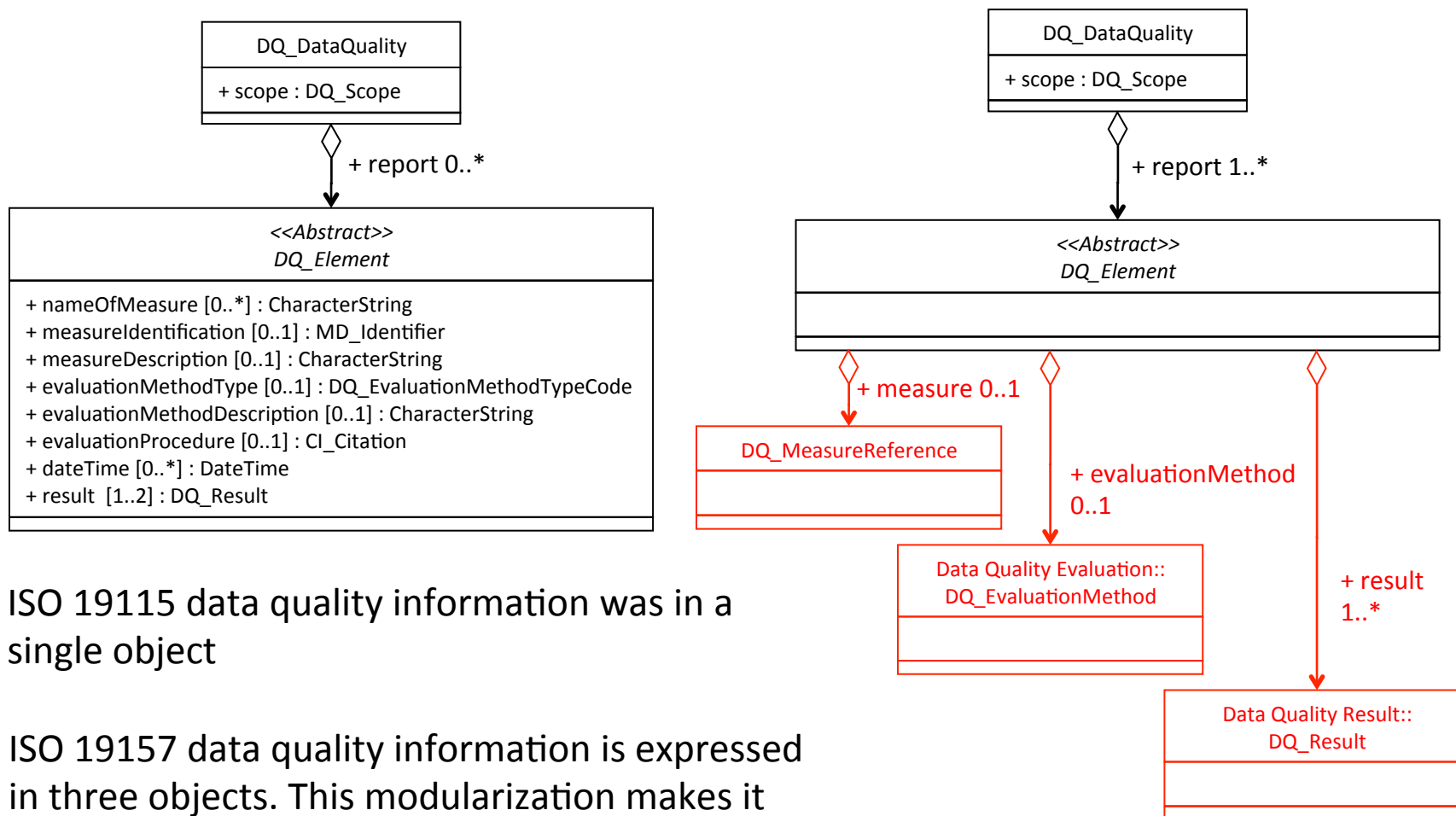|			
+ attribute	+ feature	+ initiative	+ metadata
+ attributeType	+ featureType	+ stereomate	+ initiative
+ collectionHardware	+ propertyType	+ sensor	+ sample
+ collectionSession	+ fieldSession	+ sensorSeries	+ document
+ dataset	+ software	+ platformSeries	+ repository
+ series	+ service	+ productionSeries	+ aggregate
+ nonGeographicDataset	+ model	+ transferAggregate	+ product
+ dimensionGroup	+ tile	+ otherAggregate	+ collection
			+ coverage
			+ application

Standalone Quality Reports

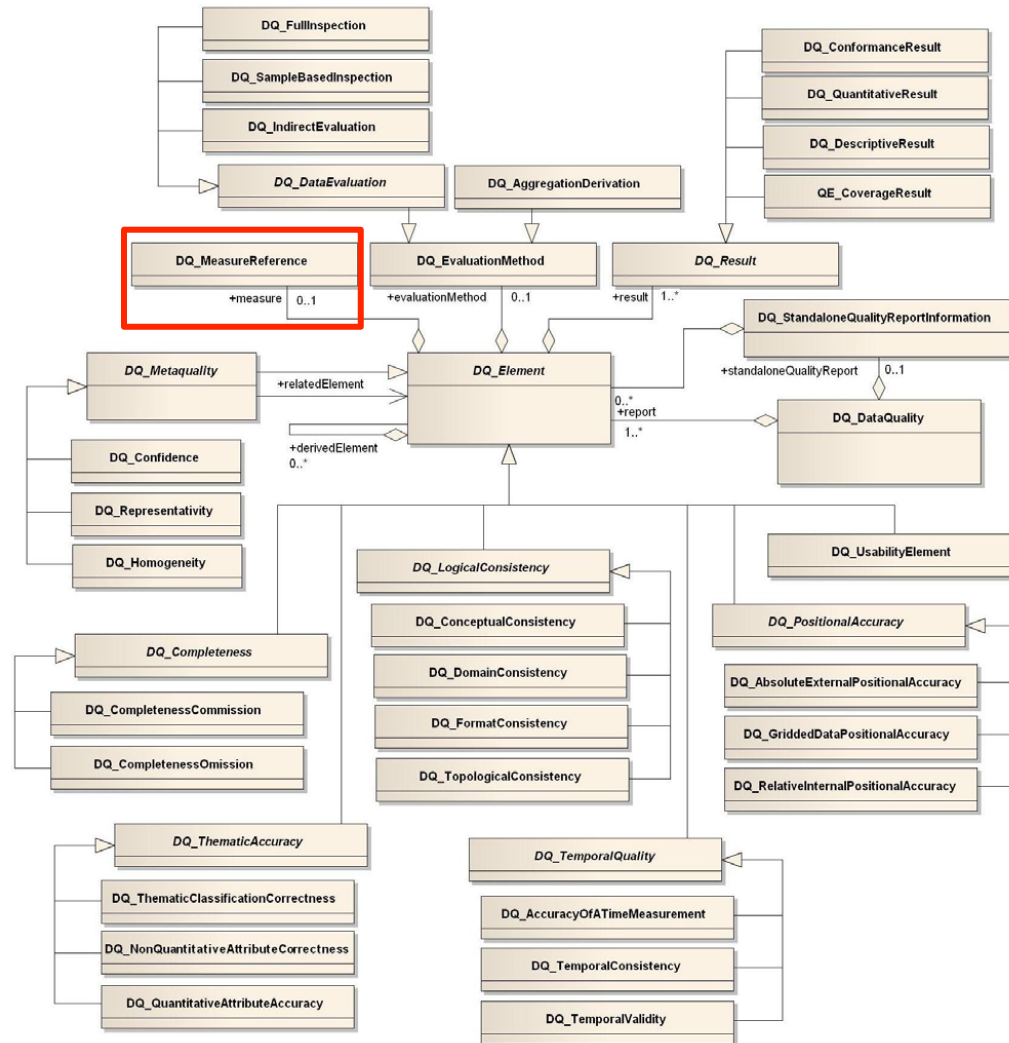ISO 19115 data quality included any number of reports that were described using the ISO 19115 conceptual model

ISO 19157 data requires at least one report (could be nil) and acknowledges that important quality information can exist in standalone reports that may not fit easily into the ISO conceptual model. These standalone reports are cited with abstracts that summarize the results for the user without neccessitating access to the cited report.

Modular Quality Information

DQ_DataQuality
+ scope : DQ_Scope

+ report 0..*

<<Abstract>>
DQ_Element

+ nameOfMeasure [0..*] : CharacterString
+ measureIdentification [0..1] : MD_Identifier
+ measureDescription [0..1] : CharacterString
+ evaluationMethodType [0..1] : DQ_EvaluationMethodTypeCode
+ evaluationMethodDescription [0..1] : CharacterString
+ evaluationProcedure [0..1] : CI_Citation
+ dateTime [0..*] : DateTime
+ result [1..2] : DQ_Result

DQ_DataQuality
+ scope : DQ_Scope

+ report 1..*

<<Abstract>>
DQ_Element

+ measure 0..1

DQ_MeasureReference

+ evaluationMethod
0..1

Data Quality Evaluation::
DQ_EvaluationMethod

+ result
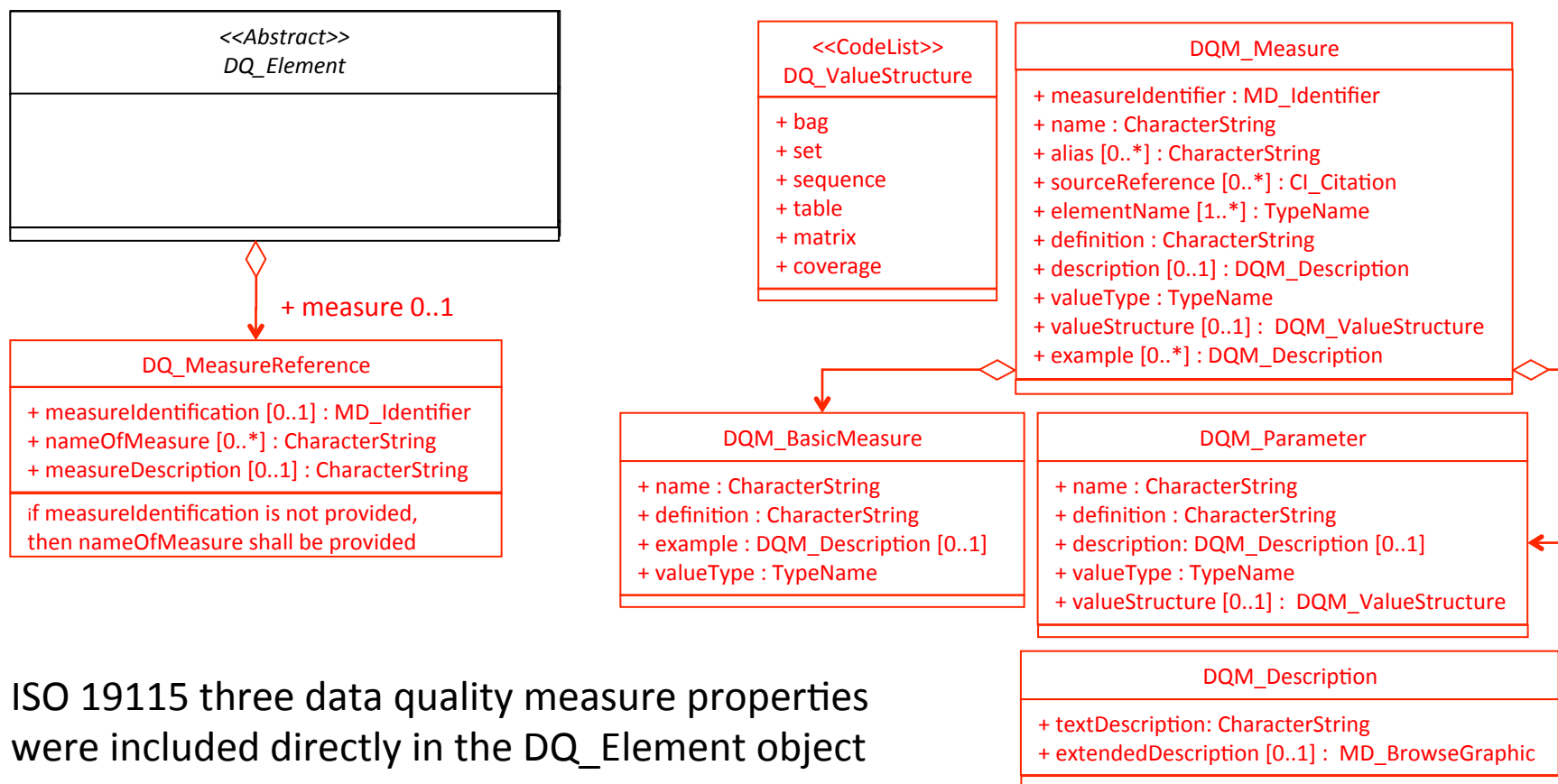1..*

Data Quality Result::
DQ_Result

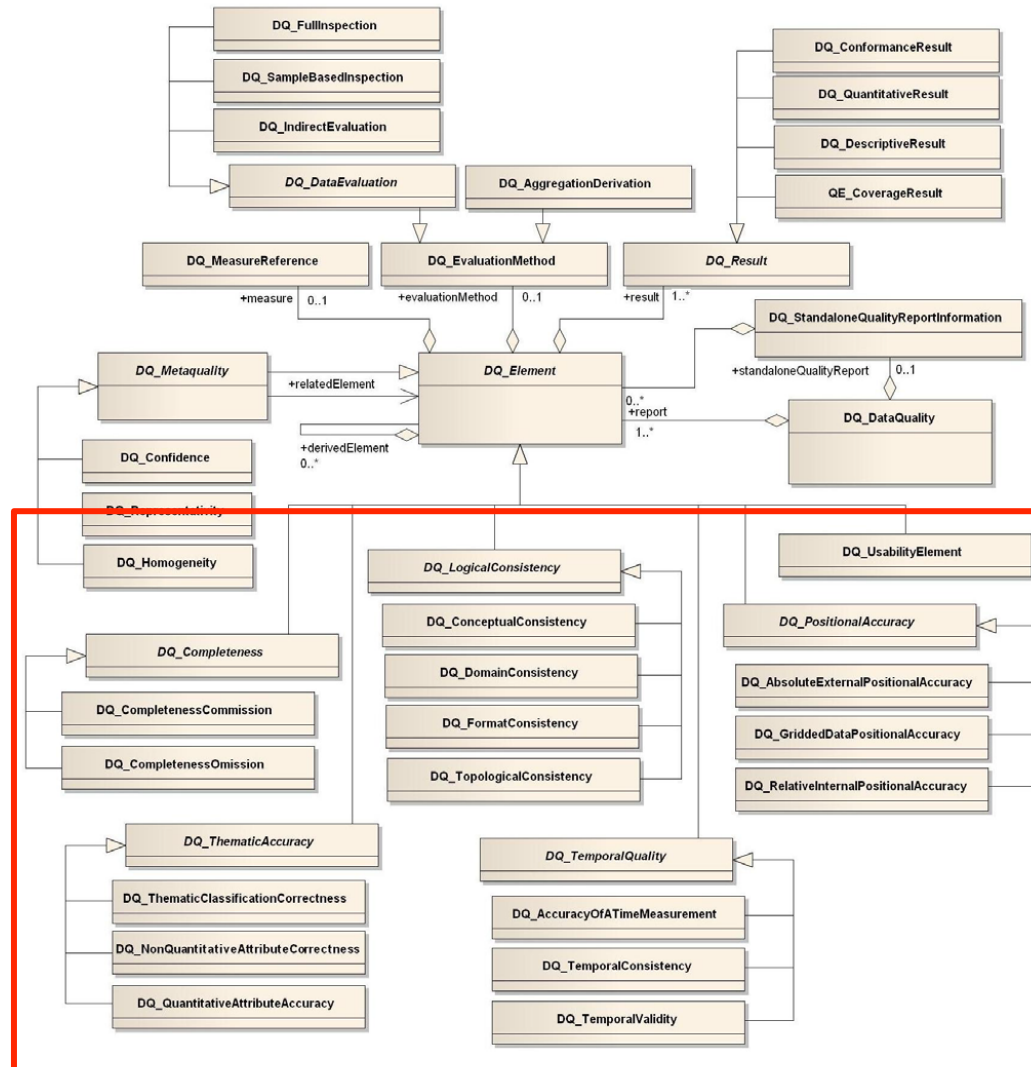ISO 19115 data quality information was in a single object

ISO 19157 data quality information is expressed in three objects. This modularization makes it easier and more flexible to connect to existing data quality systems.

<<Abstract>>
DQ_Element

+ measure 0..1

DQ_MeasureReference

+ measureIdentification [0..1] : MD_Identifier
+ nameOfMeasure [0..*] : CharacterString
+ measureDescription [0..1] : CharacterString

if measureIdentification is not provided,
then nameOfMeasure shall be provided

**<<CodeList>>
DQ_ValueStructure**

+ bag
+ set
+ sequence
+ table
+ matrix
+ coverage

DQM_Measure

+ measureIdentifier : MD_Identifier
+ name : CharacterString
+ alias [0..*] : CharacterString
+ sourceReference [0..*] : CI_Citation
+ elementName [1..*] : TypeName
+ definition : CharacterString
+ description [0..1] : DQM_Description
+ valueType : TypeName
+ valueStructure [0..1] : DQM_ValueStructure
+ example [0..*] : DQM_Description

DQM_BasicMeasure

+ name : CharacterString
+ definition : CharacterString
+ example : DQM_Description [0..1]
+ valueType : TypeName

DQM_Parameter

+ name : CharacterString
+ definition : CharacterString
+ description: DQM_Description [0..1]
+ valueType : TypeName
+ valueStructure [0..1] : DQM_ValueStructure

DQM_Description

+ textDescription: CharacterString
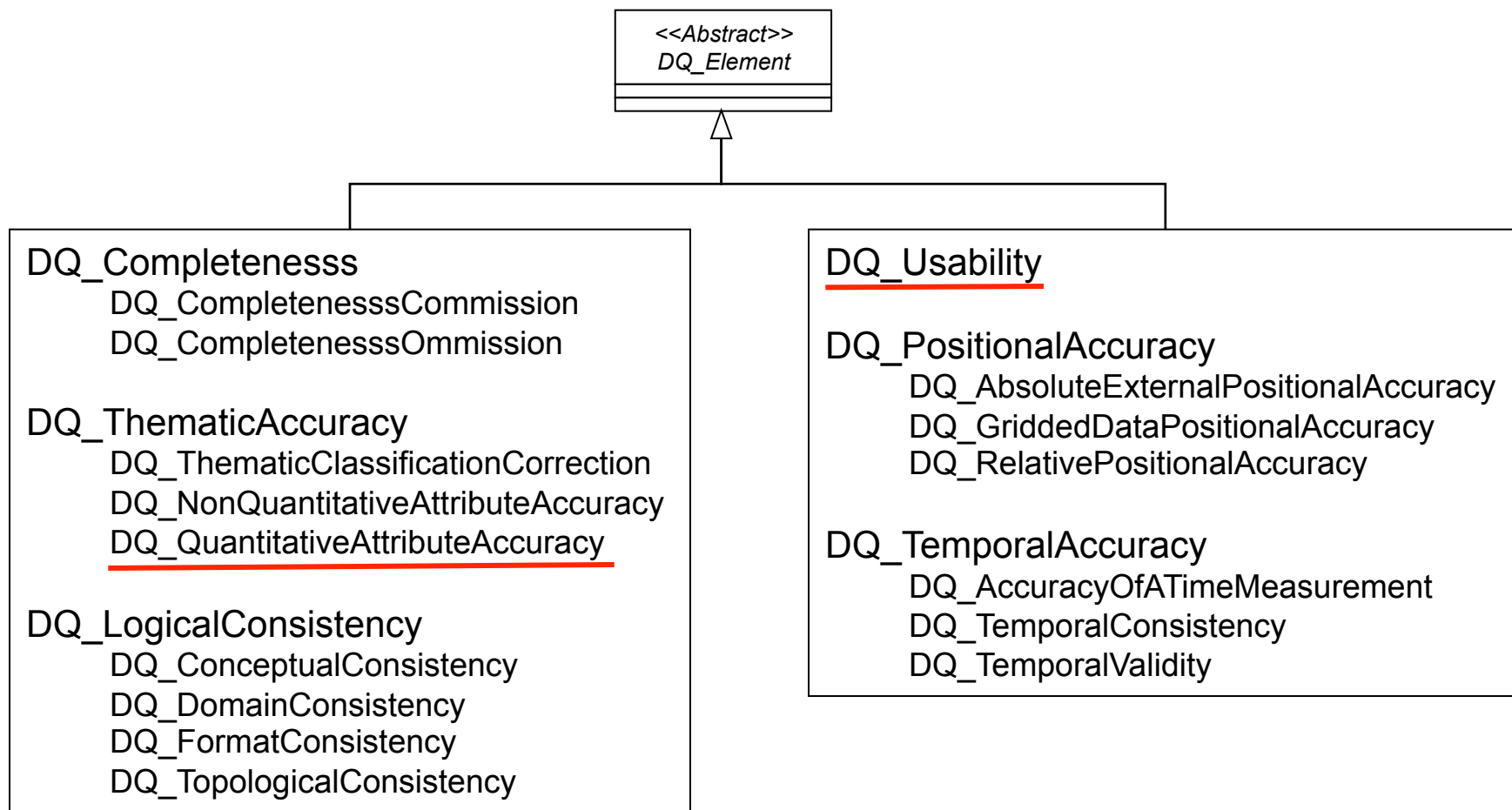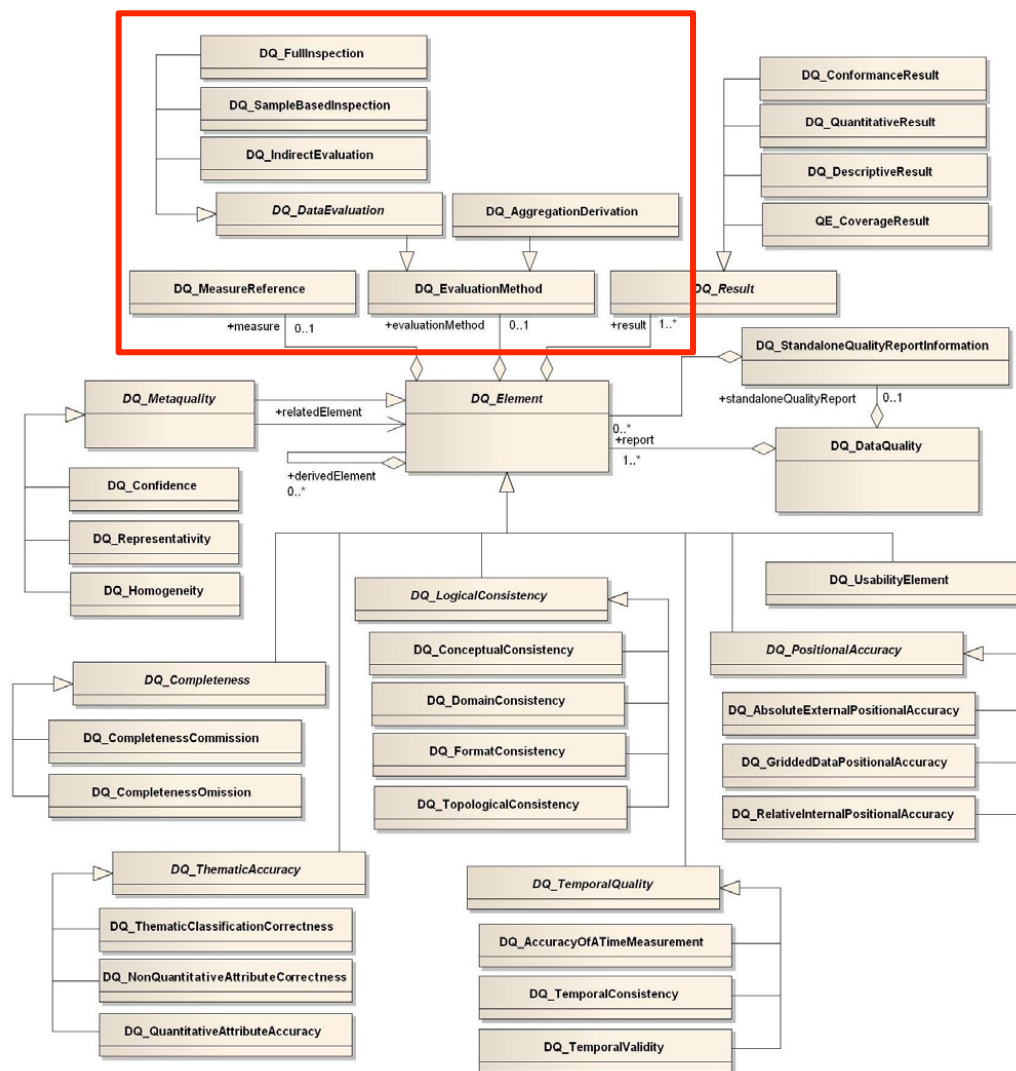+ extendedDescription [0..1] : MD_BrowseGraphic

ISO 19115 three data quality measure properties
were included directly in the DQ_Element object
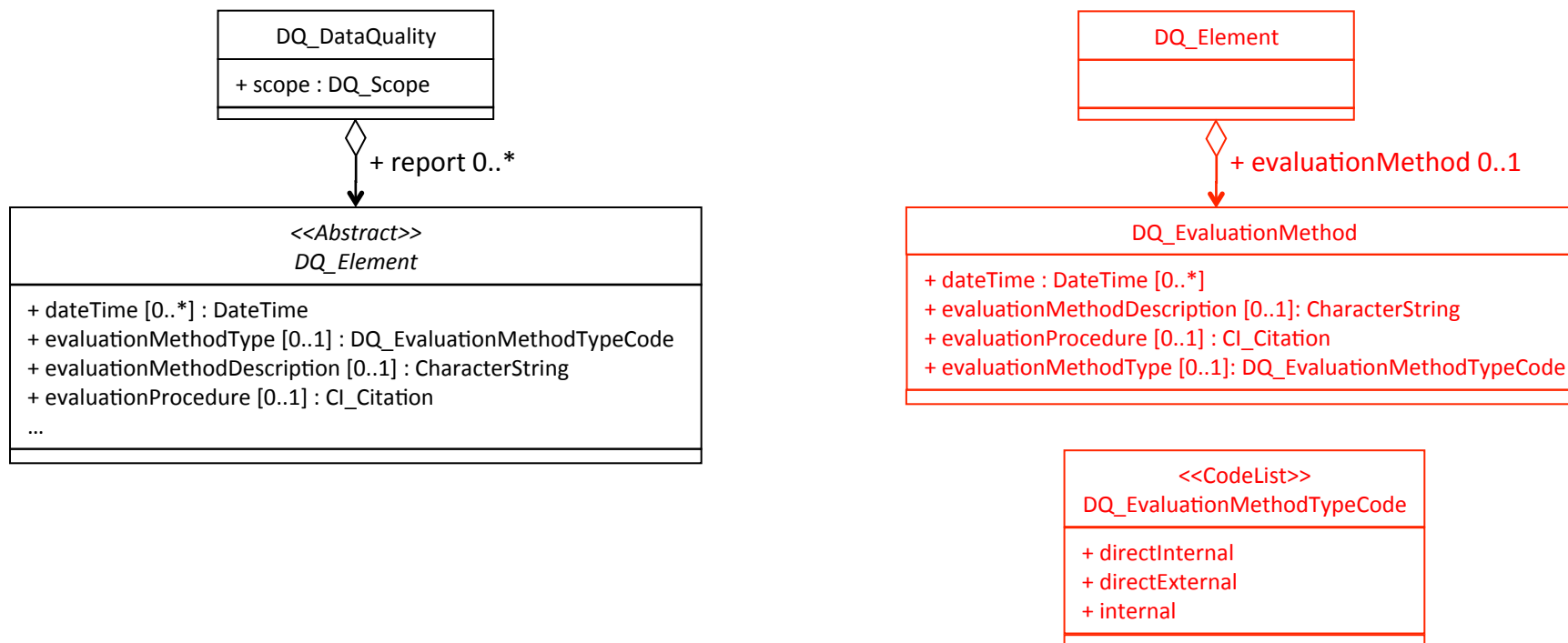
ISO 19157 data quality measure properties are in a
separate object and provide a much more complete
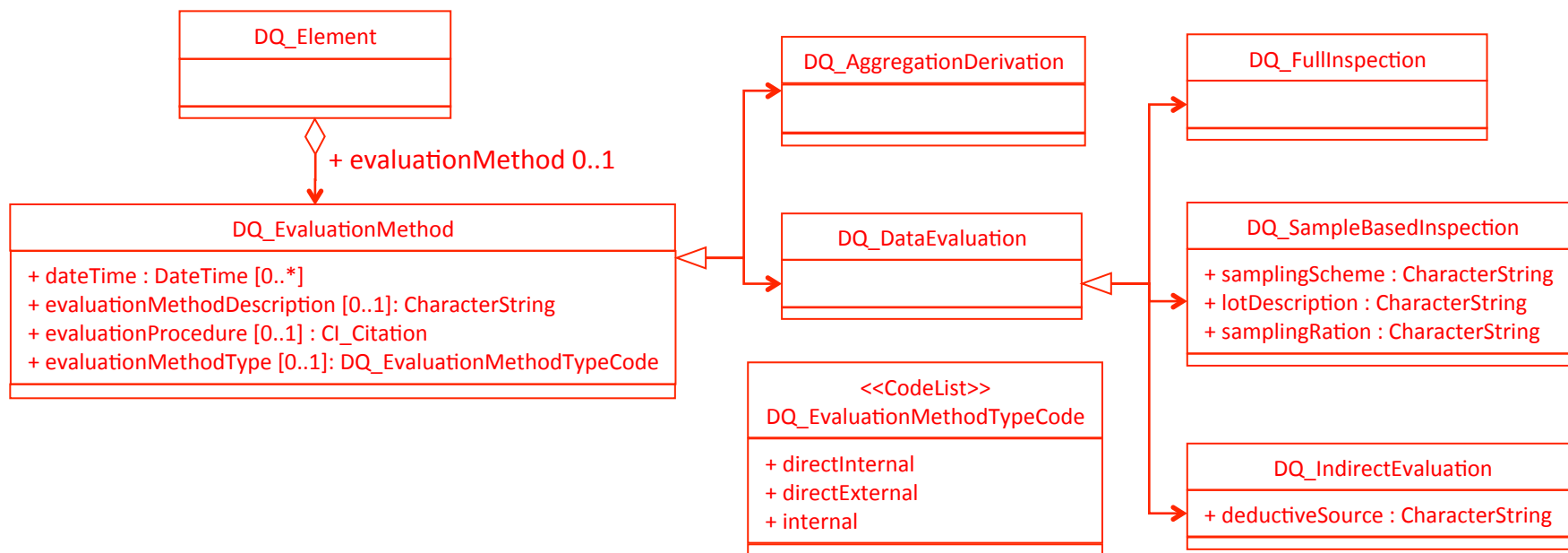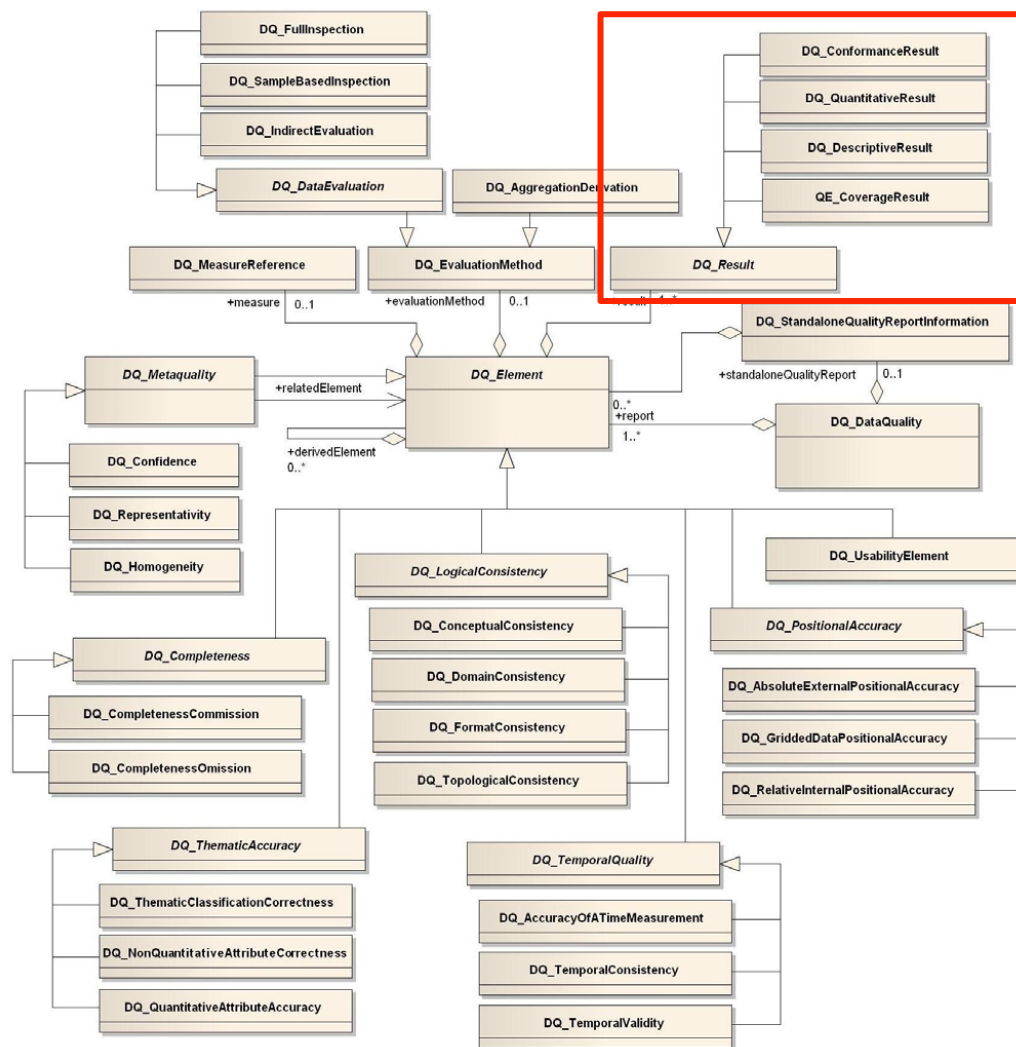description of the measures.

Data Quality Element Types

<<Abstract>>
DQ_Element

DQ_Completenesss
    DQ_CompletenesssCommission
    DQ_CompletenesssOmmission

DQ_ThematicAccuracy
    DQ_ThematicClassificationCorrection
    DQ_NonQuantitativeAttributeAccuracy
    DQ_QuantitativeAttributeAccuracy

DQ_LogicalConsistency
    DQ_ConceptualConsistency
    DQ_DomainConsistency
    DQ_FormatConsistency
    DQ_TopologicalConsistency

DQ_Usability

DQ_PositionalAccuracy
    DQ_AbsoluteExternalPositionalAccuracy
    DQ_GriddedDataPositionalAccuracy
    DQ_RelativePositionalAccuracy

DQ_TemporalAccuracy
    DQ_AccuracyOfATimeMeasurement
    DQ_TemporalConsistency
    DQ_TemporalValidity

Did not change between 19115 and 19157

Data Quality Evaluation Method

ISO 19115 three data quality evaluation properties were included directly in the DQ_Element object
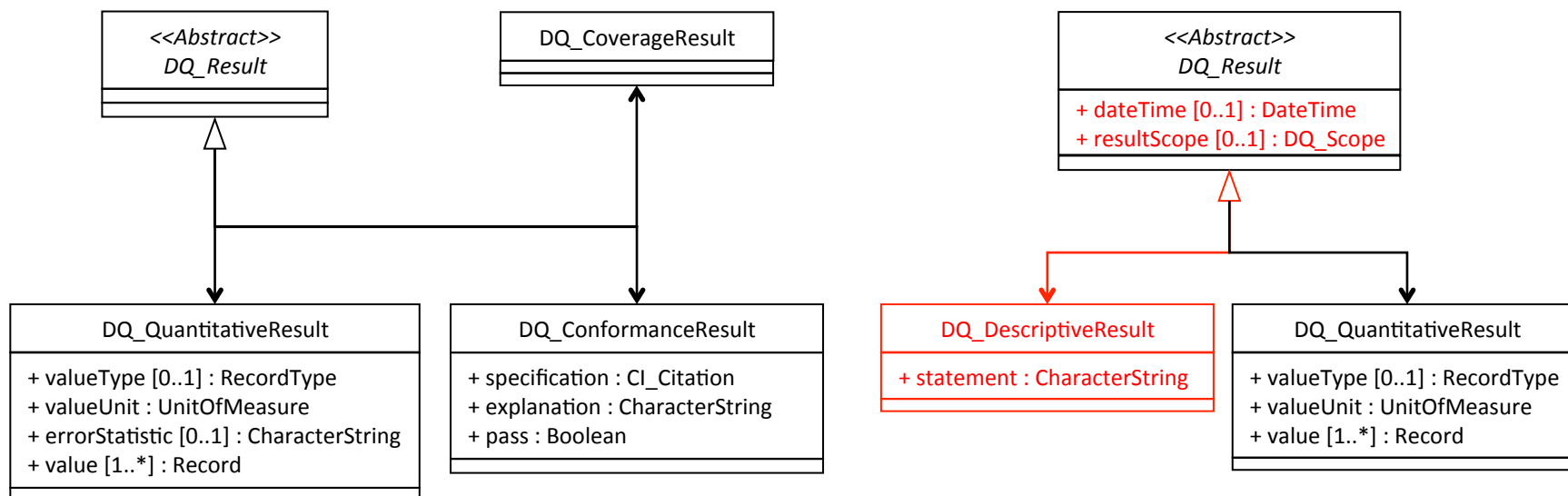
ISO 19157 data quality evaluation properties are in a separate object.

ISO 19115 allowed the specification of the evaluation as internal or external.
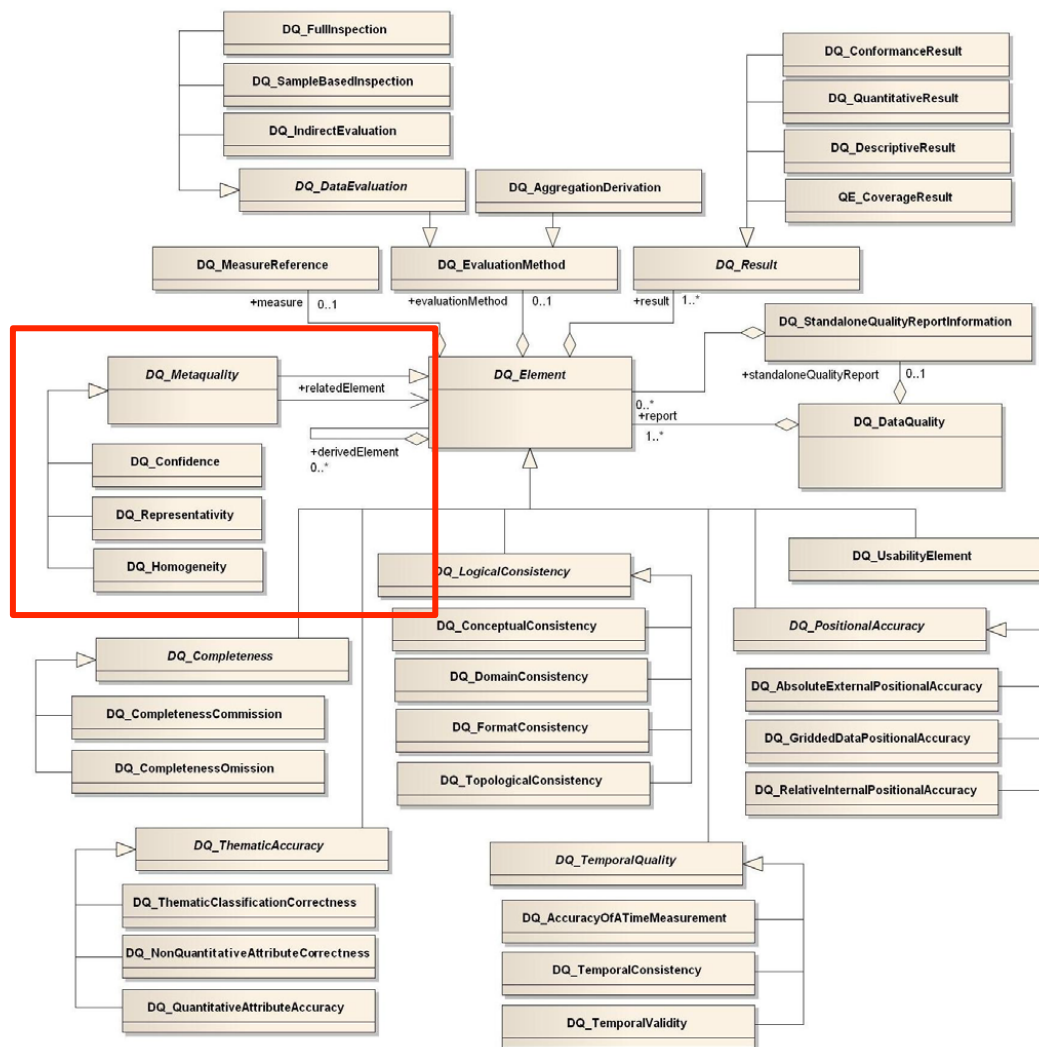
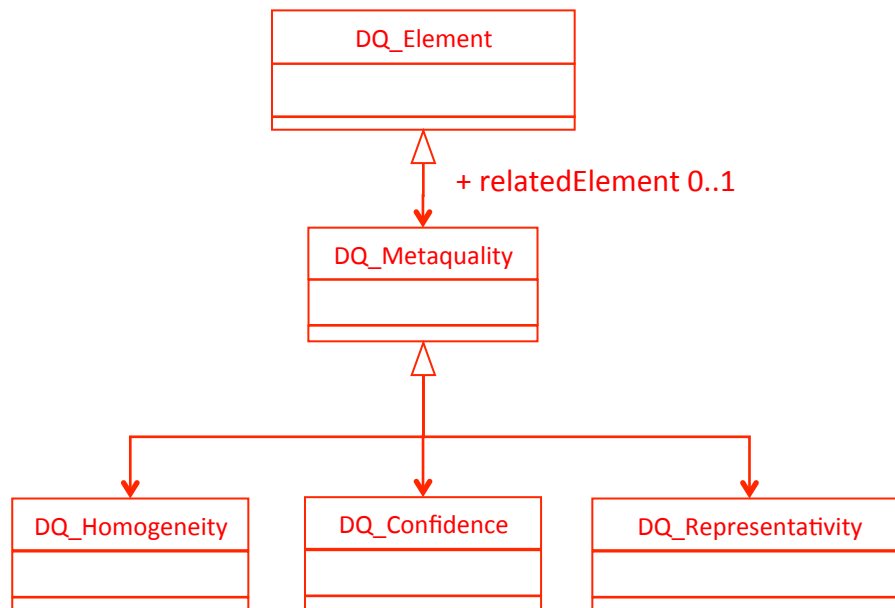ISO 19157 provides more details about these evaluation techniques.
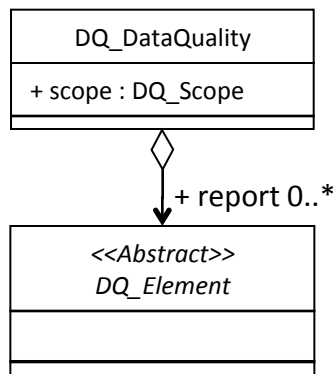
Data Quality Results

ISO 19115(-2) included three kinds of data quality reports: DQ_QuantitativeResult, DQ_ConformanceResult, and DQ_CoverageResult (from 19115-2).
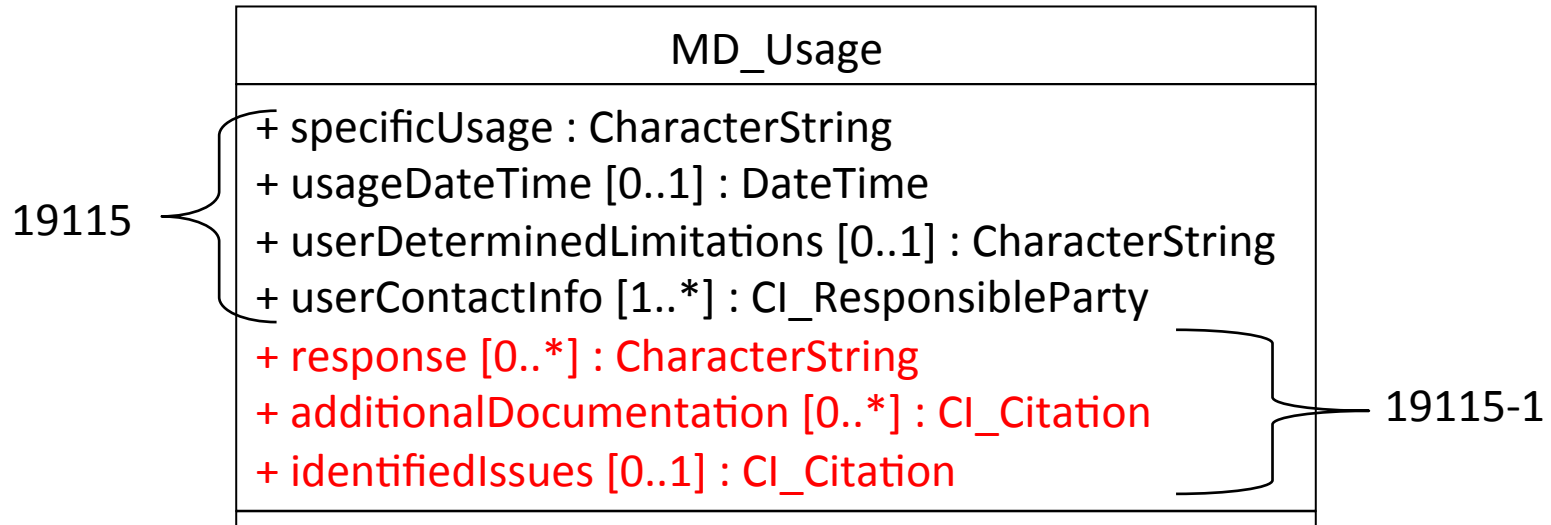
ISO 19157 adds a resultScope that allows multiple scopes in a single DQ_DataQuality object and includes a new kind of report (DQ_DescriptiveResult) that includes a simple text description of the result of the quality test.

Metaquality Information

Metaquality Information

ISO 19115 provided a complete model for reporting quality, but no ability to describe the quality of the quality determination

ISO 19157 includes three types of metadata quality reports.

```
                    MD_Usage

        + specificUsage : CharacterString
        + usageDateTime [0..1] : DateTime
 19115  + userDeterminedLimitations [0..1] : CharacterString
        + userContactInfo [1..*] : CI_ResponsibleParty
        + response [0..*] : CharacterString
        + additionalDocumentation [0..*] : CI_Citation        19115-1
        + identifiedIssues [0..1] : CI_Citation
```
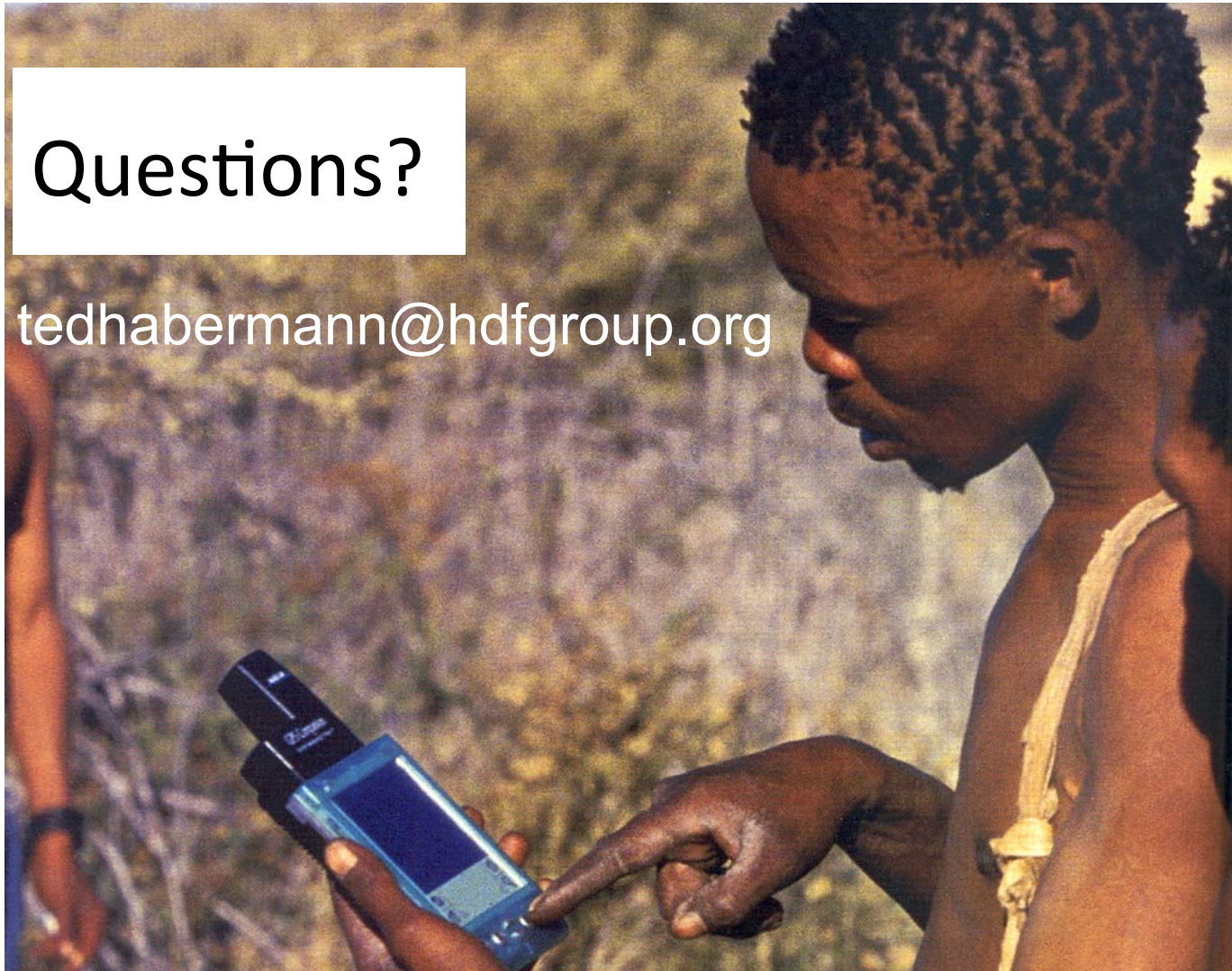
Data are used in a variety of sometimes unexpected ways. In some cases, those users identify limitations or usability problems. ISO 19115 includes information about when and how users are using data as well as limitations that they identify.

ISO 19115-1 introduces the capability for data providers to document known limitations and include responses to those users. This is an important mechanism for helping users understand how datasets are improved or updated in response to their input.

Questions?

tedhabermann@hdfgroup.org

Acknowledgements